试论大数据时代数据新闻记者的数据意识及其培养

刘永喜
（开封日报社， 河南 开封 475000）

摘　要：大数据时代下信息量激增，记者通过对这些海量数据的挖掘、分析和创新性呈现，推动了数据新闻的兴起。但在数据新闻的实践中，记者由于缺乏数据意识，在数据新闻的来源把关、采写传播、呈现方式及价值理念等方面面临巨大挑战。为了促进数据新闻的进一步发展，记者应尽快树立起理性的数据意识：既要加强数据处理和新技术应用能力，提高数据素养，又要认识到数据的人为性，复杂性，避免陷入数据陷阱。

关键词：大数据；数据新闻；把关；人文素养；采写；可视化　　**中图分类号：**G647　　**文献标识码：**A
文章编号：1671-0134（2021）01-105-03　　**DOI：**10.19483/j.cnki.11-4653/n.2021.01.032
本文著录格式：刘永喜．试论大数据时代数据新闻记者的数据意识及其培养 [J]. 中国传媒科技，2021（01）：105-107.

导语

数据新闻是基于数据分析和计算机技术的可视化新闻样式，在新闻叙事中使用数据呈现原本仅靠文字所难以呈现的内容；或者通过数据分析发现问题，并进而挖掘出新闻故事。[1]作为大数据时代正在兴起的一种跨学科、跨领域的新闻报道方式，目前数据新闻的选题集中在政治（主要是选举活动）、财经、能源和环境等与公共事务和国计民生密切相关，但又不容易通过文字或图表等传统手段理解和阐明的领域。[2]据悉，早在2009年包括《卫报》《纽约时报》《华盛顿邮报》、BBC 及 CNN 等在内的多家国际知名媒体就开展了与数据新闻相关的业务。2012 年，在谷歌新闻实验室（Google News Lab）和奈特基金会（Knight Foundation）的支持下，众多媒体联合在全球范围内设置了首个为嘉奖数据新闻领域杰出作品的奖项：数据新闻奖（DataJournalism Awards， DJA）。相较于数据新闻在国外学界和业界研究的如火如荼，我国关于数据新闻的学理探讨和实践运用开展的相对较晚。2011 年，我国的三大门户网站搜狐、新浪、网易分别开设了“数字之道”“图解新闻”“数读”等栏目[3]，拉开了数据新闻在中国实践的序幕。随后不久，新华社、人民日报和《壹读》等媒体也开始了数据新闻的探索之路。

当下，数据成为承载现实社会变动和人类的思维和情感的载体，其中蕴含的大量新闻线索，成为记者获得新闻信息、创造新闻价值的源泉。[4]由此带来的记者工作模式和价值理念的巨大变革使记者的工作被赋予了新的内涵：以可视化作为新闻主要的呈现方式，以数据处理分析的结果作为驱动报道逻辑的核心。记者的扮演的角色也更加多元：不仅仅要对信息的收集、设计和生产负责，也要参与到内容的推广流程中去。即记者的工作已经从最初的选题设计贯穿到最终用户反馈的整个过程中。

1. 数据新闻记者面临的问题

大数据时代下，从整体上看，目前数据新闻中记者面临的挑战主要表现在来源把关、制作与传播、呈现方式及价值理念四方面。

1.1　数据的来源把关

在数据新闻的实践中，新闻来源的改变，为当下记者的信息收集采写工作带来了两大方面的挑战。首先，数据来源广泛，信息繁杂。其真实性不易核实且缺乏权威性。其次，目前我国记者可获得的数据来源包括由政府机构、社会组织、企业等采集并公开的数据、产生于各类媒体平台上的用户数据、移动终端上的地理信息、物联网上的各种状态数据等。[5]但由于网络的匿名特性，个人、组织、机构等在提供数据时的目的动机、负责程度以及精确度等方面存在差异，导致数据在可信度方面也有差别：既有信度较高的数据，也有虚假失真的数据，甚至还有恶意作假的不良数据[6]，加上数据库本身对数据不假思索的囊括，其中必然包含大量未经分析核实的数据。其次，新闻管控使数据公开有限且相关法律并未完善，记者获取数据的渠道并不多，一定程度上限制了数据新闻记者的数据信息采集范围。最后，个人信息保护机制不健全，记者有时从不规范渠道获取个人数据信息，一定程度上侵犯了公众隐私。

1.2　新闻的采写传播

记者对数据的收集、调查和分析程度是决定新闻报道水平质量的基础，更是评判记者能否在纷繁复杂的社会表象中找到主线和问题的关键。目前，记者对数据的深度挖掘和加工程度还处于一个较低的水平，即便是表面上看罗列了很多图表和数据，但并不能凭借数据进行深度报道。其次，我国新闻记者还未形成系统的数据意识，对多维度数据的有序整合、排列、呈现等再结构化方式

不熟悉甚至陌生，导致在数据分析中主观色彩浓厚，把随机化结果当做规律性结论、把相关关系认为是因果关系[7]，造成报道偏颇，有失客观中立等问题。如 2017 年 DT 财经网发表的《五道口生存指南：浪在“宇宙中心”是怎样一种体验？》一文，记者仅仅抓取了五道口周边的签到微博评论，就贸然对该数据进行文本及词频分析，其最终得出的“从整体看来，五道口附近的人们生活过得还是比较开心”[8] 的结论明显缺乏因果逻辑：在五道口附近微博签到的用户不一定居住在此，更重要的是，一瞬间的心情也无法代表整个的生活常态。该新闻片面的观点不可避免地引发了公众的不休争论。

1.3 交互的呈现方式

可视化的交互是数据新闻的一大亮点：通过将原本无序、杂乱和晦涩的数据转换为既有逻辑可循又能有效互动的新闻报道，不仅能调动用户的阅读兴趣，还可以增强用户的阅读体验。现实情况是，一方面，记者往往会忽视数据新闻的适用范围，滥用可视化形式阐释新闻，增加公众理解新闻的难度。另一方面，记者对搜集到的冗杂繁琐的数据信息的筛选、过滤和分析流于表象，使新闻作品缺乏吸引力和交互性，影响新闻报道的传播效果。比如，网易数读在 2018 年推出的题为《2018 第一张汪星人全家福，祝您狗年大吉吧》的新闻中，记者以狗为主体共搜集了 87 张插图，个人认为繁杂且单一的图片不仅影响读者的阅读体验，更干扰了读者对主题信息的接收。

1.4 记者的价值理念

数据新闻对数据技术的推崇和依赖使新闻记者面临人性和技术性的失衡困境。在传统新闻的观念体系中，新闻的主体是人，新闻的主旨是服务于人，新闻的叙事更是围绕个体的存在、生命、故事展开。但目前我国记者在数据新闻的实践中，过于侧重从数据关系中发现社会问题，从而展开相对宏观的社会探讨，对个体的情感体验，价值认知的阐释不到位。这种重视数据新闻的普遍性，忽视个性、注重可视化技术的运用，忽视新闻的叙事角度的行为，严重影响公众的认知体验和认同方式。[9]

2. 数据新闻中记者数据意识的培养

在大数据时代，记者既不能忽视数据在新闻报道中越来越重要的作用，也不能为了追求数据，忽视长期积累的人文素养、对人情世态的洞察力对新闻报道的重要性。因此，为了推动数据新闻更好地发展，记者应尽快树立起正确的数据意识，从数据新闻的来源把关、采写传播、呈现方式及价值理念四个方面不断提高自身的数据素养。

2.1 加大数据新闻来源的权威度和可靠性

在媒介化社会中，记者对数据的收集、审查及分析能否在短时间内让读者掌握关键信息的新闻报道显得十分重要。首先，记者要对数据来源进行翔实的调查并清晰标注来源。若数据过于繁琐、冗杂和无序，可以联合多个平台记者共同建构数据来源追溯机制。当仅靠记者自身不易对数据来源进行审查核实时，也可以在新闻中给出数据来源的链接或文件，调动公众去质疑和求证数据的真实可靠性。据统计，2017 年数据新闻奖获奖作品中，有 89% 作品都清晰标注数据来源，其中 46% 的作品以链接的形式提供了原始数据的获取方式。其次，面对目前我国的数据大部分来源于各类政府部门、企业单位和研究机构等第三方，数据公开制度不完善的现实情况，记者要在向用户开放和分享数据的基础上，采取合理规范的方式吸引用户在自己的平台上创建数据，从而建立自己的数据库。最后，记者要合理合法使用数据，避免侵犯公众的数据隐私。

2.2 提高数据新闻采写的科学性和专业性

大数据环境改变了记者的工作流程和职能定位，记者不仅要有基本的专业知识和选题策划能力，也要充分发挥自己对社会的敏锐感知力，加大数据新闻报道的传播度和影响力。首先，记者可以根据自身对公众需求和社会问题的独特见解，从细微生活入手，发现新闻价值，策划新闻选题。同时，借助数据技术手段，挖掘数据背后的相关性，搭建个体到社会的桥梁。比如 DT 财经推行的数据新闻项目“地铁 1 公里”。DT 财经以地铁站为基准，收集分析了周边 1 公里区域各类大数据，主要包括房产中介机构（居住数据）、高德地图（交通数据、公交数据和高德指数）、大众点评（商家 POI 数据）和云房数据研究中心等数据。在综合交叉不同数据后，为“地铁 1 公里”建立了一个包含服务配套、餐饮娱乐、交通人气、房价、办公功能等多项指标的综合评价体系，作为观察中国城市各区域人口流动、商业变化和经济发展的生态轨迹。整个报道既富有人情味和趣味性，又具有深刻的社会价值。其次，因社交媒体不仅可以快速，广泛地传播新闻，还可以成为数据新闻生产的来源。因此，记者要重视并合理运用社交媒体的数据聚合和发散功能：一方面，通过对社交平台上的用户进行信息抓取和分析；另一方面，鼓励社交平台上的用户参与到新闻的线索提供与内容传播中。比如《卫报》记者曾根据总统大选结果向全球用户征询意见，根据收到的反馈数据，制作了《奥巴马再次当选美国总统：全球民众的态度》报道。最后，就是要掌握基本的数据分析和数据挖掘方法，利用数据技术和逻辑思维解读社会问题，不让技术成为从传统新闻向数据新闻转型成功的瓶颈。

2.3 增强数据新闻呈现方式的可视化和简洁化

记者对数据处理的结果呈现是否简洁美观、是否有

趣生动、是否符合当下人们碎片化和及时性的阅读习惯，是做好可视化数据新闻的关键。首先，在理念上，记者要清楚地认识到数据新闻不等于使用了数据的新闻，也不是指由数据构成的新闻，它具有一定的操作适用性，并非所有的新闻都适合做数据新闻。数据只有在使用的过程中才能转化为信息，并通过与故事结合呈现其意义。记者在新闻工作中，应根据新闻类型、内容及对象，选择合适、清晰、准确的形式搜集、挖掘及分析数据，帮助公众更好地理解新闻传达的核心内涵。其次，在前期工作中，记者要重视数据采写的量化研究方法和研究思路，尽可能找到或是挑选出适合新闻报道题材类型的、能与用户有效互动的、能提升阅读体验的数据信息。如《武汉保卫战的关键信息都在这张图上》一文通过妥善运用数据，明晰地呈现出 29 个省市对武汉的对口支援状况：以武汉市地图为背景，以 29 个箭头代表各省市的医护援助。同时，将“武汉”二字用心形标识凸显，彰显出我国在面对灾难时“一方有难，八方支援”的团结景象，更隐喻了政府和人民坚决打赢这场防疫战的决心和信念。[10] 这一新闻报道的可视化和简洁化都是值得借鉴的。

2.4 坚持数据新闻价值理念的人文性和公众性

数据本身的意义是有限的，数据新闻记者使用数据的目的不仅是为了预测未来或是找出数据背后隐藏的社会模式，更是为了挖掘数据背后人与人的关系及意义。因此，在新闻数据采编制作中要加强思辨意识和批判精神，提高自身的人文性素养，体现公共关怀和同情心。如南方周末特稿《中国环境病人》一文讲述的是居住在污染源周边，饱受“公害病”却没能得到法律保护的居民现状。报道以个体赵光明为突破口，论证了这个新闻事件这并不是一个人的悲剧，而是社会的悲剧：38 年间，磺厂医院有 400 名癌症患者离世，其中 90% 属砷中毒。[11] 这里的数据使用，不只是为了呈现事实，更是落脚在了由个体生命到深层社会议题的审视，引发了公众对环境保护、弱势群体生存等问题的热烈讨论，达到了传统文字报道难以达到的传播效果。

结语

数据新闻不只是简单的事实呈现，优秀的数据新闻还可以提供深刻的观点，搭建从个体到社会的桥梁，揭示数据背后所隐含的生命价值。记者通过对冗杂无序的数据分析，可以深入挖掘数据背后的相关性；用清晰、易读和可视化的呈现方式，阐释复杂多变的局面，达到纯文字报道难以达到这样的传播效果。与此同时，记者要理性地看待数据。新闻故事的生动性和深度性远非数据所能代替，长期的轻量化阅读会影响大众思辨能力的提高 [12]，进而阻碍社会发展。因此，数据新闻记者一方面要重视数据在新闻中的应用，加快树立起正确的数据意识：既要有数据敏感，准确又迅速地捕捉到具有新闻价值的数据信息；又要积极学习视觉设计工具，培养数理统计分析能力，不断提升自身的数据素养。另一方面，数据新闻记者也要警惕“唯数据主义”的倾向，坚守人文精神和新闻理想，透过数据积极凸显其背后的社会价值和个体生命内涵，帮助人们在不确定的世界中寻找到复杂问题的解决方案。媒

参考文献

[1] 吴小坤 . 数据新闻：理论承递、概念适用与界定维度 [J]. 新闻与传播研究，2017（10）：120–126.
[2] 李希光，张小娅 . 大数据时代的新闻学 [J]. 新闻传播，2013（01）：7–11.
[3] 袁满，强月新 . 我国数据新闻研究的回顾与前瞻 [J]. 郑州大学学报，2016（2）：130–135.
[4] 李祥甫，李光庆 .5G 时代做好数据新闻及可视化的思考 [J]. 城市党报研究，2020（09）：87–88+91.
[5][10] 文卫华，李冰 . 大数据时代的数据新闻报道——以英国《卫报》为例 [J]. 现代传播，2013（05）：139–142.
[6] 丁柏铨 . 数据新闻：价值与局限 [J]. 编辑之友，2014（07）：6–10.
[7] 沈浩，谈和，文蕾 . 数据新闻记者的数据素养 [J]. 青年记者，2014（21）：17–18.
[8] 罗婕，邓绍根 . 数据新闻生产传播中的问题及应对 [J]. 青年记者，2017（22）：41–42
[9] 刘涛 . 理解数据新闻的观念：可视化实践批评与数据新闻的人文观念反思 [J]. 新闻与写作，2019（04）：65–71.
[11] 郁轩，肖涛 . 人民日报微信公众号数据新闻的可视化叙事——以新冠肺炎疫情报道为例 [J]. 新闻研究导刊，2020（16）：235–236.
[12] 张军辉 . 从“数字化”到“数据化”：数据新闻叙事模式解构与重构 [J]. 中国出版，2016（08）：39–43.

作者简介：刘永喜（1976–），男，河南开封，研究生，新闻中级，研究方向：新闻业务。

（责任编辑：胡杨）